Evaluating Index-based Treatment Allocation in Underresourced Communities

Niclas Boehmer[*1,2], **Yash Nair**[*3], **Sanket Shah**[*1], **Lucas Janson**[1], **Aparna Taneja**[4], **Milind Tambe**[1,4]
[1]Harvard University
[2]Hasso Plattner Institute, University of Potsdam
[3]Stanford University
[4]Google Deepmind
niclas.boehmer@hpi.de, yashnair@stanford.edu, sanketshah@g.harvard.edu, ljanson@fas.harvard.edu,
aparnataneja@google.com, milind_tambe@harvard.edu

Abstract

In many applications of AI for Social Impact (e.g., when allocating spots in support programs for underserved communities), resources are scarce and an allocation policy is needed to decide who receives a resource. Before being deployed at scale, a rigorous evaluation of an AI-powered allocation policy is vital. In this paper, we introduce the methods necessary to evaluate index-based allocation policies, which allocate a limited number of resources to those who need them the most. Such policies create dependencies between agents, rendering standard statistical tests invalid and ineffective. Addressing the arising practical and technical challenges, we describe an efficient estimator and methods for drawing valid statistical conclusions. Our extensive experiments validate our methodology in practical settings while also showcasing its statistical power. We conclude by proposing and empirically verifying extensions of our methodology that enable us to reevaluate a past randomized control trial conducted with 10 000 beneficiaries for a mHealth program for pregnant women. Our new methodology allows us to draw previously invisible conclusions when comparing two different ML allocation policies.

1 Introduction

In treatment allocation, we have a limited number of intervention resources. The challenge is to devise an allocation policy that decides who gets a resource to maximize social welfare. Treatment allocation finds applications in various domains, particularly prevalent in underresourced communities. Examples include when (i) allocating scarce medical resources such as medication (Ayer et al. 2019; Deo et al. 2013) or screening tools (Lasry et al. 2011; Deo et al. 2015; Bastani et al. 2021), (ii) scheduling inspection visits, e.g., for railways (Gerum, Altay, and Baykal-Gürsoy 2019) and wind farms (Yeter, Garbatov, and Soares 2020), or (iii) allocating spots in support programs, e.g., after-school programs (Mac Iver et al. 2019) or vaccination (Kehinde et al. 2024) and mHealth programs (Verma et al. 2023b) in underserved communities.

Treatment allocation is a common problem in statistics and economics (Kitagawa and Tetenov 2018), and AI methods are increasingly used to solve treatment allocation problems of societal importance (Killian et al. 2021, 2019; Künzel et al. 2019; Bastani et al. 2021). Two common strategies are to make allocations based on predicted individualized measures of risk (Kent et al. 2016; Mac Iver et al. 2019) or treatment effects (Künzel et al. 2019; Verma et al. 2023b). Both of these and many other strategies can be captured by *index-based allocation policies*. Given a fixed number of resources, these policies first compute an index (e.g., engagement or risk) for each individual and subsequently allocate the resources to the individuals with the lowest index. While there is a rich body of work on the design of such allocation policies (see references above), we study the orthogonal problem of *evaluating the effectiveness of a policy* using randomized control trials (RCTs) (Hariton and Locascio 2018). The ability to quantify the benefits of different index-based allocation policies is crucial to deciding which policy should be deployed in the field. This problem comes with novel technical and practical challenges: Whether or not an individual gets selected for treatment by an allocation policy depends on the other beneficiaries in the population. The resulting dependence between beneficiaries renders central assumptions behind standard statistical tests invalid (see Section 3). Unfortunately, the problem of evaluating allocation policies has received almost no attention so far and is particularly prevalent in social domains, where field trials have limited participants and are expensive to conduct.

Bridging this gap, *we provide the necessary tools to effectively draw reliable statistical conclusions about the quality of index-based allocation policies*. We describe a new estimator together with customized statistical inference techniques and demonstrate its impact in the context of the large mobile health program mMitra (Murthy et al. 2020) run by the Indian NGO ARMMAN in which an index-based allocation policy was deployed. This program provides critical preventive health care information to enrolled pregnant women and mothers of infants from underserved communities. To promote engagement in the program, health workers can call a limited number of beneficiaries each week to provide them with additional information and guidance. The deployed policy for allocating these service calls has been developed and refined over a series of papers (Mate et al. 2022; Verma et al. 2023b; Wang et al. 2023; Shah et al. 2024) and has impacted more than 350K beneficiaries since deployment.

To establish the benefit of live service calls and inform policy design, Mate et al. (2022) and Verma et al. (2023a)

[*]These authors contributed equally.

conducted RCTs to evaluate the effectiveness of index-based allocation policies based on different ML paradigms. They use an intuitive standard estimator (Section 3), to which we refer as the base estimator. Unfortunately, the estimator lacks flexibility and suffers from low statistical power (see Section 5), necessitating the repetition of RCTs due to small effect sizes and noise. Moreover, Mate et al. (2022) and Verma et al. (2023a) note that their methodology comes without empirical evidence or theoretical guarantees on the validity of computed confidence intervals and drawn statistical conclusions, limiting the impact of their analysis and evaluation. Addressing this challenge, in Section 4, we present the *subgroup estimator*, which computes the average treatment effect by comparing those who were selected by the policy in the policy arm of the RCT to those the policy *would have selected* in the control arm. While the base estimator requires the execution of customized RCTs using the analyzed policy, our methodology is also compatible with standard RCTs where treatment is allocated uniformly at random. Translating ideas from Imai and Li (2024), we prove that the subgroup and base estimator are asymptotically normal and describe methods for computing asymptotically valid confidence intervals for evaluating and comparing policies.

In Section 5, we use synthetic and real-world data to build simulators. We successfully verify that our asymptotic theoretical guarantees regarding the validity of confidence intervals for our estimators empirically extend in various setups and computed confidence intervals remain valid for as few as 500 agents or budgets of 100. Moreover, we demonstrate that the subgroup estimator typically has a significantly higher statistical power (up to a factor of 8) than the base estimator, which allows us to draw many previously hidden conclusions.

In Section 6, we turn to a field trial conducted for the mMitra program (Verma et al. 2023a), which has been so far only evaluated using the natural, yet unvalidated, base estimator. This has questioned the reliability of the previously drawn conclusions regarding the usefulness of the deployed resource allocation method for mMitra. Reevaluating the field trial comes with additional practical challenges for our methodology, e.g., accounting for the sequential allocation of resources, measuring long-term intervention effects, and dealing with imbalanced covariates. Addressing these challenges, we present extensions of our methodology and use them to reevaluate the field trial. For the first time, we provide reliable conclusions relevant to the NGO, e.g., we present the long-sought proof for the benefit of their deployed machine learning method for allocating interventions on improving beneficiaries' engagement in the program. ARMMAN has already used our evaluation methodology in a recent study on establishing the positive influence of AI-prescribed interventions on health outcomes (Dasgupta et al. 2024) and plans to use it in future RCTs. Additional details and results can be found in our full version (Boehmer et al. 2024).

2 Preliminaries

Let $A = \{0, 1\}$ be the set of actions, where 1 is the active action (treatment/intervention is given) and 0 is the passive action (no treatment given). An agent $i \in [n]$ is characterized by covariates $\mathbf{x}_i \in \mathcal{X}$ and a reward function $R_i : A \to \mathbb{R}$ that returns the reward generated by the agent given the action assigned to it. Agents are drawn i.i.d. from a probability distribution P defined over the space of covariates and reward functions $\mathcal{X} \times (A \to \mathbb{R})$. If not stated otherwise, expectations and probabilities are over groups of n agents sampled i.i.d. from the probability distribution P. In this case, we write $\mathbf{X}_n := (\mathbf{x}_i)_{i\in[n]}$ to denote the covariates of these n agents and $(R_i)_{i\in[n]}$ for their reward functions.

An allocation policy π gets as input the covariates $\mathbf{X}_n \in \mathcal{X}^n$ of n agents and a treatment fraction α and returns $\lceil \alpha n \rceil$ agents receiving the active action (alternatively, we could specify the *number* of agents receiving a treatment). We denote as $J_i^{\pi(\mathbf{X}_n,\alpha)}$ the indicator variable that denotes whether agent $i \in [n]$ gets assigned a treatment by policy π, i.e., $J_i^{\pi(\mathbf{X}_n,\alpha)} = 1$ if $i \in \pi(\mathbf{X}_n, \alpha)$. An index-based allocation policy π^Υ is defined by a function $\Upsilon : \mathcal{X} \to \mathbb{R}$ mapping covariates to an index. Given $\mathbf{X}_n \in \mathcal{X}^n$ and a treatment fraction $\alpha \in [0, 1]$, π^Υ returns the $\lceil \alpha n \rceil$ agents $i \in [n]$ with the lowest index $\Upsilon(\mathbf{x}_i)$. Further, given $\mathbf{X}_n \in \mathcal{X}^n$ and a threshold $\lambda \in \mathbb{R}$, let $\upsilon^\Upsilon(\mathbf{X}_n, \lambda)$ return the agents $i \in [n]$ with an index value $\Upsilon(\mathbf{x}_i)$ smaller or equal to λ. We refer to the policy that acts on everyone in υ^Υ as a *threshold policy* (which notably does *not* satisfy the definition of an allocation policy, as the number of agents that receive an active action is not fixed).

RCT Design In previous RCTs (Mate et al. 2022; Verma et al. 2023a), treatment is allocated according to the evaluated policy: We have access to the results of an RCT with a policy arm (p) containing n agents $(\mathbf{x}_i^p, R_i^p)_{i\in[n]}$ sampled i.i.d. from P on which we run our policy π. As the outcome, we observe $(\mathbf{x}_i^p, R_i^p(J_i^p))_{i\in[n]}$, where $J_i^p := J_i^{\pi(\mathbf{X}_n^p,\alpha)}$. Moreover, we have access to a control arm (c) of n agents $(\mathbf{x}_i^c, R_i^c)_{i\in[n]}$ sampled i.i.d. from P for which we observe $(\mathbf{x}_i^c, R_i^c(0))_{i\in[n]}$. Note that for both the control and policy arm, we naturally can only observe the agent's reward according to the action applied to them, while the counterfactual remains unobserved.

Statistics Notation An estimand is the quantity we want to measure, and an estimator is a value that "approximates" the estimand, computed from the available observed data. Estimands' names will always involve a τ, while estimators' names will always involve a θ. A sequence of random variables $(A_n)_{n>0}$ with cumulative distribution functions $(G_n(a))_{n>0}$ *converges in distribution* to a random variable A with cumulative distribution function G if $\lim_{n\to\infty} G_n(a) = G(a)$ for all $a \in \mathbb{R}$ at which G is continuous, in which case we write $A_n \xrightarrow{d} A$ (for us, n will typically be the number of samples we observe, i.e., the number of people in the RCT). We denote as $\mathcal{N}(\mu, \sigma^2)$ the normal distribution with mean μ and variance σ^2. Let q_α be the smallest number so that an expected α-fraction of agents has an index below q_α.

3 Challenges and Previous Work

We describe previous approaches for evaluating index-based allocation policies, their shortcomings, and the work of Imai and Li (2024) on which many of our ideas are built.

Previous Approach The approach taken in previous work (Mate et al. 2022; Verma et al. 2023a; Mate et al. 2023) was

to estimate the average benefit an agent derives from being a member of the policy arm instead of the control arm. This translates to measuring the difference between the expected reward generated by an (arbitrary) agent from the policy and control arm: $\tau^{\text{base}}_{n,\alpha}(\pi) = \frac{1}{n}(\mathbb{E}\sum_{i\in[n]} R_i(J_i^{\pi(\mathbf{X}_n,\alpha)}) - \mathbb{E}\sum_{i\in[n]} R_i(0))$. To estimate $\tau^{\text{base}}_{n,\alpha}(\pi)$, they compute the difference in the observed generated reward of all agents in the policy arm compared to all agents in the control arm:

$$\tilde{\theta}^{\text{base}}_{n,\alpha}(\pi) = \frac{1}{n}\Big(\sum_{i\in[n]} R_i^p(J_i^p) - \sum_{i\in[n]} R_i^c(0)\Big) \qquad (1)$$

To make this estimator consistent with our subgroup estimator (Section 4.1), we rescale $\tilde{\theta}^{\text{base}}_{n,\alpha}$ and let $\theta^{\text{base}}_{n,\alpha}(\pi) := \frac{n}{\lceil \alpha n \rceil}\tilde{\theta}^{\text{base}}_{n,\alpha}(\pi)$ be the *base estimator*. The work of Mate et al. (2023) tries to improve upon the base estimator; however, their methods do not allow for the computation of confidence intervals that are necessary for hypothesis testing. In our full version, we argue that our subgroup estimator works similarly but comes with (valid) confidence intervals. Other previous works on evaluating index-based allocation policies, e.g., the works by Perdomo (2024) and Shirali, Abebe, and Hardt (2024), either focus on different evaluation aspects or make strong assumptions on the environment.

The Challenge of Dependent Samples The base estimator θ^{base} has two main shortcomings: It suffers from low statistical power, i.e., the estimator is quite "noisy" leading to large confidence intervals and problems in distinguishing between policies (see Section 5). Further, while this estimator seems intuitive, there are no theoretical guarantees or empirical evidence that computed confidence intervals and drawn statistical conclusions are valid (as acknowledged by Mate et al. (2022) and Verma et al. (2023a)).

To understand these problems, let us consider the class of threshold policies (Section 2), which make independent decisions for every individual. Standard statistical inference methods, which rely on the central limit theorem (CLT), can be used for these threshold policies. The CLT says that the sample mean of independent observations drawn from some (arbitrary) distribution (as generated, e.g., by a threshold policy in an RCT) converges to a normal distribution. Estimates of this normal distribution's mean μ and variance σ^2 can then be used for estimating the variance of the estimator and, for instance, to construct valid confidence intervals. However, for resource allocation policies, the samples we observe in the policy arm are no longer independent because an agent's treatment and, thereby, their observed reward depends on the index values of other agents. This renders the standard central limit theorem inapplicable. Consequently, statistical tests such as Welch's z-test, which rely on the CLT, are no longer guaranteed to produce accurate statistical conclusions. Thus, we face the challenge of computing valid confidence intervals and p-values for policy evaluation.

Another consequence of the dependence between agents is that if we apply an allocation policy to a group of n agents, we only observe a *single* independent group sample; slightly changing the composition of the group could change the treatment allocation and thereby also the observed rewards (in contrast, for threshold policies, we would derive n fully independent samples). Due to the resulting lack of independent samples, we face the challenge of constructing estimators that do not suffer from low statistical power (see Section 5).

Imai and Li (2024) We discuss recent work by Imai and Li (2024), which does not make any explicit connections to allocation policies and treatment allocation. Instead, it is positioned in the growing body of work on estimating conditional heterogeneous average treatment effects (CATEs) of individuals based on their covariates (Wager and Athey 2018; Künzel et al. 2019; Kennedy 2023), applicable, for instance, when making decisions on patients in precision medicine. In fact, upon a closer inspection of this literature, only the work of Imai and Li (2024) is closely related to our problem, as they are the only ones that consider the average treatment effects in *groups of agents*. While this problem is not directly connected to allocation policies and their general results cannot be directly applied to our setting, we will use one of their lemmas making minimal assumptions on the used index function to establish theoretical guarantees for estimators for index-based allocation policies.

Imai and Li (2024) analyze how to estimate the average treatment effect in groups of agents with similar CATEs. They assume access to a standard RCT where everyone in the treatment arm receives treatment, a setup that might be infeasible if resources are scarce. Their methodology (adapted to our setting) aims to estimate the average effect a treatment has on agents with an index value below q_α, i.e., those agents who belong to the expected α-fraction of agents with the lowest index: $\tau^{\text{q}}_\alpha(\Upsilon) := \mathbb{E}_{(\mathbf{x},R)\sim P}[R(1) - R(0) \mid \Upsilon(\mathbf{x}) \leq q_\alpha]$. Alternatively, τ^{q}_α can be interpreted as quantifying the effectiveness of a policy that acts on everyone who has an index below q_α. Note that this is *not* an allocation policy because it does not act on a fixed number of agents. To measure this estimand, they take the difference between the summed reward of the α-fraction of agents in the treatment arm with the lowest indices and the summed reward of the α-fraction of agents in the control arm with the lowest indices as the estimator: $\frac{1}{\lceil \alpha n \rceil}\Big(\sum_{i\in\pi^\Upsilon(\mathbf{X}^p_n,\alpha)} R_i^p(1) - \sum_{i\in\pi^\Upsilon(\mathbf{X}^c_n,\alpha)} R_i^c(0)\Big)$. In Lemma S2 appearing in Appendix S3 of Imai and Li (2024) they show that this estimator converges to the estimand $\tau^{\text{q}}_\alpha(\Upsilon)$ at a $\sqrt{n}$-rate, a result that will enable us to use their results to establish guarantees in our setting.

4 Methodology

4.1 Subgroup Estimator

We propose a new estimand that quantifies the effectiveness of a policy by measuring the average effect of *one* treatment as prescribed by the policy. This is in contrast to the base estimand τ^{base} that measures the average effect of being an agent in the treatment group (regardless of whether or not you are treated). More concretely, for an allocation policy π, a treatment fraction α, and a group size $n \in \mathbb{N}$, we define $\tau^{\text{new}}_{n,\alpha}(\pi)$ to be the expected additional reward generated by an intervention allocated according to policy π:

$$\tau^{\text{new}}_{n,\alpha}(\pi) := \frac{1}{\lceil \alpha n \rceil}\mathbb{E}\sum_{i\in\pi(\mathbf{X}_n,\alpha)} (R_i(1) - R_i(0)) \qquad (2)$$

$\tau^{\text{new}}_{n,\alpha}(\pi)$ is, up to rescaling, equivalent to $\tau^{\text{base}}_{n,\alpha}(\pi)$:

$$\tau^{\text{base}}_{n,\alpha}(\pi) = \frac{1}{n}\Big(\mathbb{E}[\sum_{i\in[n]} R_i(J_i^{\pi(\mathbf{X}_n,\alpha)}) - \sum_{i\in[n]} R_i(0)]\Big)$$

$$= \frac{1}{n}\,\mathbb{E}\,[\sum_{i\in\pi(\mathbf{X}_n,\alpha)}(R_i(1)-R_i(0)) + \sum_{i\notin\pi(\mathbf{X}_n,\alpha)}(R_i(0)-R_i(0))] = \frac{\lceil\alpha n\rceil}{n}\tau^{\text{new}}_{n,\alpha}(\pi)$$

The reason for this equivalence is that—in expectation—agents on which we did not act in the policy arm cancel out with the corresponding agents in the control arm. Nevertheless, in the base estimator θ^{base} (which is τ^{base} after dropping the expectations), these agents will introduce noise, as they will unequally influence the observed summed reward of the two arms, i.e., the two sums in θ^{base} (cf. Equation (1)). This motivates us to "remove" them from the estimation. The *subgroup estimator* allows us to do this. We separately estimate the expected reward of agents selected by the policy when treated and when not treated. For the first part, we can use the agents selected by our policy in the policy arm (for which we observe $R_i(1)$), and for the second part, the agents that *would have been* selected by our policy in the control arm (for which we observe $R_i(0)$). This results in the subgroup estimator θ^{SG} which is equivalent to the estimator used by Imai and Li (2024):

$$\theta^{\text{SG}}_{n,\alpha}(\pi) = \frac{1}{\lceil\alpha n\rceil}\Big(\sum_{i\in\pi(\mathbf{X}^p_n,\alpha)} R^p_i(1) - \sum_{i\in\pi(\mathbf{X}^c_n,\alpha)} R^c_i(0)\Big) \quad (3)$$

In fact, it is easy to see that the expected value of the subgroup estimator $\mathbb{E}[\theta^{\text{SG}}_{n,\alpha}(\pi)]$ is equal to our estimand $\tau^{\text{new}}_{n,\alpha}(\pi)$.

Intuitive Differences between Estimators The base estimator θ^{base} treats the RCT arms as indecomposable units and estimates the effect of treatments (allocated according to policy π) on *the complete policy arm* through a comparison with the complete control arm. The idea of the subgroup estimator θ^{SG} is to estimate the effect of treatments *on the treated agents* by approximating their unobserved counterfactual behavior (when they did not get treatment) using the control arm. For this, we view each agent as an individual sample and compare the agents that received treatment in the policy arm to those that would have received treatment under the policy in the control arm. Thus, in contrast to the base estimator θ^{base}, the subgroup estimator θ^{SG} only takes into account the agents "relevant" to our policy. Specifically, θ^{SG} ignores the difference $\sum_{i\notin\pi(\mathbf{X}^p_n,\alpha)} R^p_i(0) - \sum_{i\notin\pi(\mathbf{X}^c_n,\alpha)} R^c_i(0)$ that does not provide us with any insights regarding the effectiveness of the policy and only adds noise to the estimator.

Base, Subgroup, and Hybrid Estimator The subgroup estimator has a significantly lower variance than the base estimator in our experiments from Section 5. In our full version, we explain that there are corner cases where the situation is reversed and present a hybrid estimator that combines the two, thereby blending their strengths.

Flexible RCT Design Notably, the subgroup estimator θ^{SG} offers a flexible approach to RCT design. For instance, the subgroup estimator enables us to run a standard RCT in which everyone in the treatment arm is treated and only afterward specify the index policies to be evaluated. This is in contrast to the base estimator θ^{base} that can only be used with the specific RCT design described in Section 2.

4.2 Causal Inference for Subgroup Estimator

An estimator that is often close to the estimand is insufficient for valid causal inference. What is missing is a method to quantify the estimator's error, e.g., in the form of confidence intervals, which allows one to draw high-probability conclusions. This section addresses this gap and provides informal results and ideas on how we can do asymptotically correct inference for the subgroup estimator θ^{SG} (see our full version (Boehmer et al. 2024) for details and proofs). The main ingredient for performing inference with the subgroup estimator θ^{SG} is to establish that it is asymptotically normal with respect to our estimand τ^{new}, i.e., the difference between the estimator and estimand is distributed according to a normal distribution. This allows us, for instance, to reason about the "variance" of the estimator and, thereby, the probability that the estimator's error is above a certain threshold. To establish this result, using results from Imai and Li (2024) and further observations, the general proof idea is to first show that the subgroup estimator $\theta^{\text{SG}}_{n,\alpha}(\pi^\Upsilon)$ is asymptotically normal with respect to $\tau^{\text{q}}_\alpha(\Upsilon)$. Subsequently, one can show that $\tau^{\text{q}}_\alpha(\Upsilon)$ converges "fast" to our estimand $\tau^{\text{new}}_{n,\alpha}(\pi^\Upsilon)$ to conclude that the distribution of the estimators's error converges in distributon to a normal distribution:

Theorem 4.1 (informal)**.** *Under very mild assumptions,*

$$\sqrt{n}\big(\theta^{\text{SG}}_{n,\alpha}(\pi) - \tau^{\text{new}}_{n,\alpha}(\pi)\big)/\sqrt{\sigma^2_{SG}} \xrightarrow{d} \mathcal{N}(0,1)$$

for some $\hat{\sigma^2_{SG}}$ that can be computed from the results of an RCT in closed form.

Note that we can use Theorem 4.1 to derive a formula for asymptotically correct β-confidence interval of $\tau^{\text{new}}_{n,\alpha}(\pi)$ as:

$$I = \left[\theta^{\text{SG}}_{n,\alpha}(\pi) - Z_{1-\frac{\beta}{2}}\sqrt{\frac{\hat{\sigma}^2_{\text{SG}}}{n}},\quad \theta^{\text{SG}}_{n,\alpha}(\pi) + Z_{1-\frac{\beta}{2}}\sqrt{\frac{\hat{\sigma}^2_{\text{SG}}}{n}}\right] \quad (4)$$

where Z_γ is the γ quantile of $\mathcal{N}(0,1)$. Asymptotically correct here means that the estimand is in the confidence interval I with probability converging to $1-\beta$, i.e., $\mathbb{P}(\tau^{\text{new}}_{n,\alpha}(\pi) \in I) \to 1-\beta$. Fortunately, in our experiments, we observe that the confidence interval is approximately valid already for a limited number of samples in different realistic settings.

In our full version, we discuss how Theorem 4.1 can be used for computing p-values and comparing policies (using the confidence intervals for the difference in their effects).

Inference for Base Estimator The proof of Imai and Li (2024) cannot be applied to prove the asymptotic normality of the base estimator θ^{base}. Thus, we come up with an alternative, more generally applicable proof via empirical process theory (van der Vaart and Wellner 2023) that allows us to establish a result analogous to Theorem 4.1 for the base estimator (and the hybrid estimator featured in Section 4.1). This implies that the base estimator is asymptotically valid and can be used for valid causal inference (yet, we will see that it is inferior to the subgroup estimator in terms of its power).

Domain	Estimator					
	Base			Subgroup		
	$<$ CI	in CI	$>$ CI	$<$ CI	in CI	$>$ CI
Synthetic	0.027	0.952	0.021	0.036	0.935	0.029
TB	0.024	0.946	0.030	0.031	0.947	0.022
mMitra	0.018	0.956	0.026	0.039	0.938	0.023

(a) Fraction of times the estimand is in, below ($<$ CI), or above ($>$ CI) an estimator's 95% confidence interval (over 1000 different RCTs).

Domain	Estimator	
	Base	Subgroup
Synthetic	0.426	0.178
TB	0.778	0.293
mMitra	0.668	0.221

(b) Half-width of confidence intervals (averaged over 1000 RCTs).

Table 1: Empirical comparison of the confidence intervals of the different estimators. Both the base and subgroup estimators produce approximately valid confidence intervals; however, the subgroup estimator's confidence intervals are consistently smaller.

5 Experiments

We empirically analyze the base and subgroup estimator using the variance estimation techniques described in Section 4. We are interested in (i) checking whether estimator's confidence intervals, whose asymptotic validity was established in Section 4.2, remain valid in realistic settings, and (ii) comparing the statistical power of the estimators.

Setup We assume that each agent is modeled by a MDP (Ayer et al. 2019; Verma et al. 2023a,b). We focus on adherence settings with two states ('good'$= 1$ or 'bad'$= 0$) and two actions ('intervene'$= 1$ or 'do not intervene'$= 0$), and we obtain a reward of 1 for every timestep an agent is in the good state. Our RCT arms consist of $n = 5000$ agents, and we can intervene on 20% of them ($\alpha = 0.2$). Agents transition between states according to a transition matrix T, where an entry $T^a_{s,s'}$ specifies the probability of transitioning from state $s \in \{0, 1\}$ to $s' \in \{0, 1\}$ when taking action $a \in \{0, 1\}$. We allocate the interventions in the first step. In subsequent steps, we let agents transition between states using T^0 and collect rewards for another 9 time steps. We consider three domains, differing in their transition matrices:

Synthetic Transition probabilities are sampled randomly subject to $T^1_{s,1} - T^0_{s,1} \in [0, 0.2]$ for each state $s \in \{0, 1\}$.

Medication Adherence (TB) This domain uses real-world Tuberculosis medication adherence data from Killian et al. (2019). For each agent, the data is used to fit their transition probabilities under the passive action. We then simulate the treatment effect, i.e., $T^1_{s,1} - T^0_{s,1}$, in each state $s \in \{0, 1\}$ by sampling uniformly at random from $[0, 0.2]$.

Mobile Health (mMitra) We use real-world data from a field trial (Mate et al. 2022) to evaluate the effectiveness of service calls to improve engagements in the mHealth program mMitra. Agents' transition probabilities under the active and passive action are learned from data.

We use the classic Whittle index (Weber and Weiss 1990), the standard method to solve restless multi-armed bandits, to quantify an agent's intervention effect. We want to estimate the effectiveness of this "Whittle Index" policy, focusing on computing 95% confidence intervals. We compute our estimand τ^{new} via Monte Carlo simulation.

Validity (Table 1a) We check whether the confidence intervals produced by our estimators are valid, i.e., whether the computed 95% confidence intervals (which differ between runs of an RCT for one simulator) truly contain the estimand (which is constant for each simulator) 95% of the times. Table 1a confirms that both the base and subgroup estimators produce approximately valid confidence intervals, with the error (i.e., $|95\% - \text{in CI}|$) being less than 1.5% in all domains.

Power (Table 1b) As both estimators are valid, we can compare their power. We do this in Table 1b by comparing the (half-)width of their confidence interval. The subgroup estimator always produces tighter confidence intervals, with their width being usually *around a third* of the base estimator's confidence interval across all three domains.

A Representative Example (Figure 1) It is hard to appreciate the difference between estimators in the abstract. To make the difference more concrete, we picked one representative RCT and show in Figure 1 the confidence intervals computed by our estimators. As an example of how to read these figures, note that the fact that the confidence interval of the base estimator crosses the black vertical zero line in all three domains implies that we cannot conclude that interventions had a statistically significant positive effect using the base estimator. Figure 1 also includes the random allocation policy in red that assigns treatments uniformly at random to 20% of the agents (its confidence intervals can be correctly computed using a standard Welch's z-test). The subgroup estimator allows us to draw otherwise impossible statistical conclusions. In particular, for all three domains, based on the results of the base estimator, we cannot conclude that there is a statistically significant difference between the random and Whittle policy (their confidence intervals overlap). In contrast, for the subgroup estimator, confidence intervals for the TB and mMitra simulator are disjoint from the random ones.

Changing Hyperparameters In our full version (Boehmer et al. 2024), we analyze the influence of different hyperparameters. We vary the treatment fraction, the number of agents, observed timesteps, intervention effect (for TB and synthetic), and confidence level. The computed confidence intervals remain approximately valid for all considered variations—the error for both estimators is always less than 3% and typically around 1%. Regarding estimators' power, the subgroup estimator outperforms the base one in all considered settings. Yet, the extent varies: The difference is particularly large (up to a factor of 8) if treatment resources are scarce, there are only a few agents, agents are observed over a long period,

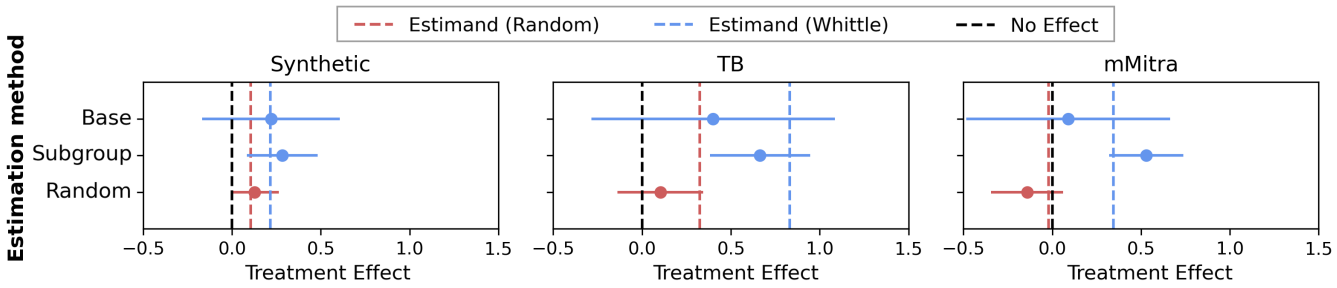

Figure 1: A representative example of the size of confidence intervals. We compare different estimators for the effectiveness of the Whittle policy (blue) and the random policy (red). The x-axis shows the average treatment effect. Vertical lines show the estimand and a zero treatment effect (black). Each estimator's point estimate is a dot, and their confidence interval is a line.

or the desired confidence level is high. The base estimator can require group sizes *ten times larger than the subgroup estimator* to achieve confidence intervals of similar size.

6 Real-World Study

We reevaluate the field trial of Verma et al. (2023a), starting by presenting different required extensions of our methodology. *This section not only demonstrates our real-world impact on the mMitra program but also provides the methodology needed in other AI4SI applications.*

6.1 Extended Methodology

We describe various extensions of our estimators addressing common real-world challenges: (i) different RCT arms might not be balanced in terms of agents' covariates, (ii) effects are measured over multiple timesteps, (iii) interventions are allocated over multiple timesteps. Our extensions are no longer covered by the techniques and guarantees from Section 4. We use the standard Welch's z-test to compute confidence intervals; we empirically verify that this produces valid confidence intervals using the same setup as in Section 5.

Covariates Mate et al. (2022) and Verma et al. (2023a) used linear regression to correct for imbalances between agents' covariates in the RCT arms common in real-world trials with limited arm sizes (Senn 2008; Kahan et al. 2014). To correct the subgroup estimator for covariates, the idea is to learn a linear function of covariates and a treatment indicator variable to capture the agent's reward. Formally, for some agent i let J_i be the action that the agent received and $x_{i,1}, \ldots, x_{i,m} \in \mathbb{R}$ be the agent's numerical covariates. We can write the regression as $R_i(J_i) = k + \beta J_i + \sum_{t=1}^{m} \gamma_t x_{i,t} + \epsilon_i$, where the coefficient β presents the average treatment effect τ^{new}. We fit the regression over the α-fraction of agents from the policy and control arm with the lowest indices, i.e., $\pi(\mathbf{X}_n^p, \alpha) \cup \pi(\mathbf{X}_n^c, \alpha)$. Note that previous work has used the agent's arm membership as the indicator variable, i.e., they replaced J_i on the right side with the agent's group membership and fitted the regression over *all* agents.

Timestep Truncation A common scenario in treatment allocation is to observe agents' behavior for T timesteps after treatments are allocated (and use their combined behavior as the reward). Choosing this T is an impactful design decision for the evaluation. If we use a small T but intervention effects last for more than T steps, we underestimate the additional reward generated by an intervention, leading to a conservative estimate. Conversely, if we pick large values of T, then the variance in agents' behavior will increase, leading to a larger variance in our estimators—decreasing T shrinks confidence intervals while simultaneously shifting them down.

Sequential Allocation In mMitra, interventions are allocated over multiple timesteps, with the constraint that each agent receives a resource in only one of them. Our subgroup estimator admits a natural extension to this setting: We take the difference between the summed reward of the agents from the policy arm that received treatment and the summed reward of the agents from the control arm that would have been allocated treatment by the policy in one of the steps.

6.2 Real-World Maternal Health Domain

We conclude by re-evaluating a real-world RCT conducted by Verma et al. (2023a). Their goal was to evaluate the effectiveness of different sequential index-based allocation policies to allocate live service calls to boost participation in ARMMAN's mMitra program (see Section 1). They follow a restless multi-armed bandits approach and use the classic Whittle index (Weber and Weiss 1990). Each RCT arm contains 3000 agents. 1800 agents are allocated service calls over 6 weeks. The reward generated by an agent is the agent's engagement in the program, i.e., the number of weeks in which they listen to a substantial part of the week's automated voice message. Verma et al. (2023a) chose to end the evaluation of their field trial after 10 weeks, i.e., the reward captures the agents' behavior for 10 weeks (including the 6 weeks where treatments are assigned); however, their data also covers the following weeks. Two index-based allocation policies are studied: "ML Method 1" is the baseline and "ML Method 2" is their improved deployed approach for index computation.[1]

Basic Results We first focus on the 10 weeks case as used in the original study (i.e., the two leftmost plots for the case with and without covariate correction). The first two rows in

[1]ML Method 1 first uses past data to learn a model for each beneficiary and then solves the allocation problem using the Whittle. In contrast, ML Method 2 follows a decision-focused learning approach (Wang et al. 2023) and combines these two steps.

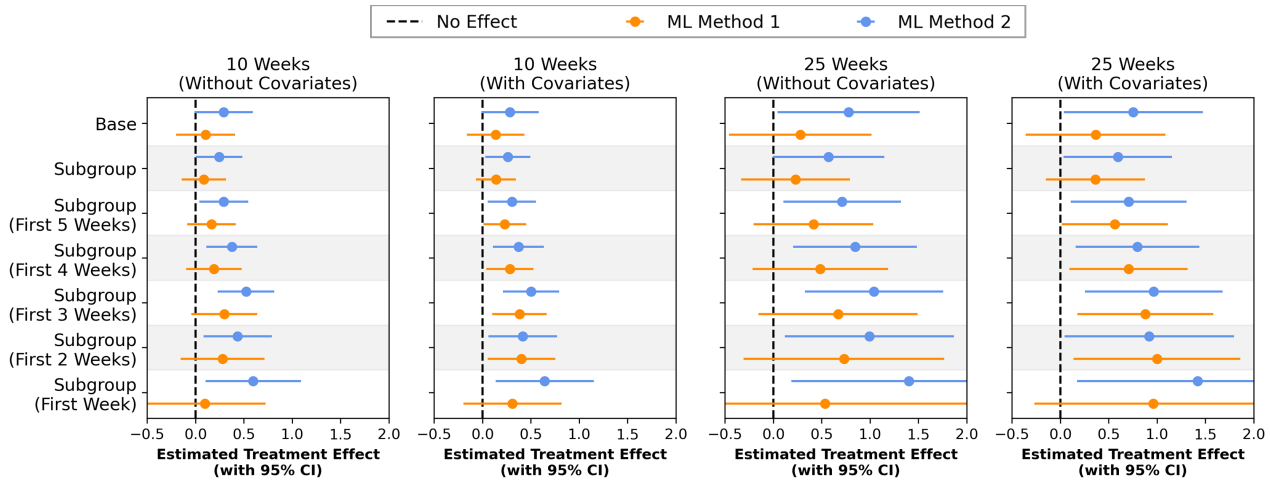

Figure 2: Evaluation of RCT from Verma et al. (2023a). We show estimators' point estimates as a dot and 95%-confidence intervals as a line for different evaluation horizons with and without correcting for covariates. "Subgroup (First x weeks)" refers to our subgroup estimator applied to all agents that (would) have been allocated a treatment up until week x.

each subfigure of Figure 2 show the results for the base and subgroup estimator. We find that using the subgroup instead of the base estimator and correcting for covariates leads to smaller confidence intervals. None of the four methods can establish a positive average effect for interventions as allocated by ML method 1 (the lower bound of the confidence interval is always smaller than 0). In comparison, for ML method 2 the lower bound for all four methods is around 0.

Fine-Grained Analysis As 60% of agents received a call throughout the trial, establishing a positive average intervention effect (on this large subpopulation) can be quite challenging: While confidence intervals tend to decrease with more agents getting treated, the effect of cleverly assigned treatments decreases in the number of allocated treatments.[2] This raises the question of whether service calls significantly positively affect at least *some* of the 1800 agents receiving them. With the previously used base estimator, answering this question would require running multiple customized trials (with varying treatment fractions and duration), as the estimator treats the policy arm as one indecomposable unit. However, the increased flexibility of the subgroup estimator allows us to address the question without rerunning the RCT. For each $x \in [1, 5]$, we can estimate the average effect of a service call on agents called in one of the first x weeks by comparing their reward to the reward of agents that would have been called in the control arm in one of the first x weeks.

Turning to the results (rows three to seven), for ML method 2, we *conclude statistically significant positive intervention effects for agents called in one of the first x weeks for $x \in [1, 5]$*, irrespective of whether we correct for covariates. Note that for $x = 1$ and no correction, we recover the single-round setting from Section 4. Thus, the conclusion that interventions—as prescribed by ML method 2—have a statistically significant effect on agents called in the first week is theoretically backed by our proofs from Section 4. *This provides proof that the NGO's interventions have a statistically significant impact on listenership for (at least) some beneficiaries identified by ML method 2.* Moreover, when correcting for covariates, service calls allocated in the first weeks by ML method 1 also have a statistically significant effect.

Enabling Smaller Budgets Verma et al. (2023a) allocated treatment to many agents because the base estimator has an enormous variance and suffers from extremely low statistical power when the treatment fraction is low. Thus, looking at the average effect on the 1800 agents is, in some sense, the best one can do with the base estimator. However, the subgroup estimator performs well even if the budget is small. As a result, it allows us to run RCTs with much lower costs, which are easier to run in underresourced communities and in which higher average intervention effects are present. Moreover, this also ensures a better alignment of the evaluation with potential real-world deployment, during which fewer resources might be available than during the RCT.

25 Week Evaluation Moving from observing beneficiaries for 10 weeks to 25 weeks has significant consequences. As featured in Section 6.1, this allows us to measure the actual long-term impact of treatment (thereby increasing the value of the estimator), while concurrently leading to (much) larger confidence intervals. However, despite this increase in the size of the confidence intervals, this leads to an increase in the lower bounds of confidence intervals here. As a result, for a confidence level of 95%, using 25 instead of 10 weeks allows us to establish up to 50% larger effect sizes, e.g., for the first three weeks for ML method 2 without covariate correction. *This enables the previously unknown conclusion that intervention effects in mMitra seem to be long-lasting.* Using our methodology to obtain such insights for other AI for Social Good programs will empower domain partners to make informed decisions toward the deployment of AI models.

[2]The high treatment fraction also partly explains the previously mentioned similarity between the two estimators, as the higher the treatment fractions, the more similar the two estimators become.

Ethics Statement

While this paper focuses on evaluating the "effectiveness" of index policies in terms of their total treatment effect, we would like to point out that before deployment index policies should also be assessed with regards to fairness and algorithmic bias aspects to ensure that they do not favor certain individuals or groups, potentially reinforcing existing inequalities.

With regards to the real-world study conducted by Verma et al. (2023a) and discussed in Section 6, we would like to remark that the used data are from fully anonymous datasets. Beneficiaries consented to data collection, and collection and sharing were done with approval by ARMMAN's ethics review committee and ethics review board. Our experiments constitute a secondary analysis of the data. Moreover, regarding the above-mentioned fairness considerations, previous research (Verma et al. 2023b, 2024) has examined the fairness of the deployed policy discussed in Section 6, demonstrating that it prioritizes the most vulnerable populations.

Acknowledgments

This work was supported by the Army Research Office (MURI W911NF1810208), by the NSF grant CBET-2112085, and by the AI Research Institutes Program funded by the National Science Foundation under AI Institute for Societal Decision Making (AI-SDM), Award No. 2229881. YN was partially supported by a Graduate Research Fellowship from the National Science Foundation. We thank Kevin Guo and James Yang for helpful discussions.

References

Ayer, T.; Zhang, C.; Bonifonte, A.; Spaulding, A. C.; and Chhatwal, J. 2019. Prioritizing hepatitis C treatment in US prisons. *Operations Research*, 67(3): 853–873.

Bastani, H.; Drakopoulos, K.; Gupta, V.; Vlachogiannis, I.; Hadjichristodoulou, C.; Lagiou, P.; Magiorkinis, G.; Paraskevis, D.; and Tsiodras, S. 2021. Efficient and targeted COVID-19 border testing via reinforcement learning. *Nature*, 599(7883): 108–113.

Boehmer, N.; Nair, Y.; Shah, S.; Janson, L.; Taneja, A.; and Tambe, M. 2024. Evaluating the Effectiveness of Index-Based Treatment Allocation. arXiv/2402.11771.

Dasgupta, A.; Boehmer, N.; Madhiwalla, N.; Hedge, A.; Wilder, B.; Tambe, M.; and Taneja, A. 2024. Preliminary Study of the Impact of AI-Based Interventions on Health and Behavioral Outcomes in Maternal Health Programs. arXiv/2402.11771.

Deo, S.; Iravani, S.; Jiang, T.; Smilowitz, K.; and Samuelson, S. 2013. Improving health outcomes through better capacity allocation in a community-based chronic care model. *Operations Research*, 61(6): 1277–1294.

Deo, S.; Rajaram, K.; Rath, S.; Karmarkar, U. S.; and Goetz, M. B. 2015. Planning for HIV screening, testing, and care at the veterans health administration. *Operations research*, 63(2): 287–304.

Gerum, P. C. L.; Altay, A.; and Baykal-Gürsoy, M. 2019. Data-driven predictive maintenance scheduling policies for railways. *Transportation Research Part C: Emerging Technologies*, 107: 137–154.

Hariton, E.; and Locascio, J. J. 2018. Randomised controlled trials—the gold standard for effectiveness research. *BJOG: An International Journal of Obstetrics and Gynaecology*, 125(13): 1716.

Imai, K.; and Li, M. L. 2024. Statistical Inference for Heterogeneous Treatment Effects Discovered by Generic Machine Learning in Randomized Experiments. *Journal of Business & Economic Statistics*, 1–13.

Kahan, B. C.; Jairath, V.; Doré, C. J.; and Morris, T. P. 2014. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials*, 15(1): 1–7.

Kehinde, O.; Abdul, R.; Afolabi, B.; Vir, P.; Namblard, C.; Mukhopadhyay, A.; and Adereni, A. 2024. Deploying ADVISER: Impact and Lessons from Using Artificial Intelligence for Child Vaccination Uptake in Nigeria. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI '24)*, 22185–22192. AAAI Press.

Kennedy, E. H. 2023. Towards optimal doubly robust estimation of heterogeneous causal effects. *Electronic Journal of Statistics*, 17(2): 3008–3049.

Kent, D. M.; Nelson, J.; Dahabreh, I. J.; Rothwell, P. M.; Altman, D. G.; and Hayward, R. A. 2016. Risk and treatment effect heterogeneity: re-analysis of individual participant data from 32 large clinical trials. *International journal of epidemiology*, 45(6): 2075–2088.

Killian, J. A.; Biswas, A.; Shah, S.; and Tambe, M. 2021. Q-Learning Lagrange Policies for Multi-Action Restless Bandits. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21)*, 871–881. ACM.

Killian, J. A.; Wilder, B.; Sharma, A.; Choudhary, V.; Dilkina, B.; and Tambe, M. 2019. Learning to prescribe interventions for tuberculosis patients using digital adherence data. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*, 2430–2438. ACM.

Kitagawa, T.; and Tetenov, A. 2018. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2): 591–616.

Künzel, S. R.; Sekhon, J. S.; Bickel, P. J.; and Yu, B. 2019. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences*, 116(10): 4156–4165.

Lasry, A.; Sansom, S. L.; Hicks, K. A.; and Uzunangelov, V. 2011. A model for allocating CDC's HIV prevention resources in the United States. *Health Care Management Science*, 14: 115–124.

Mac Iver, M. A.; Stein, M. L.; Davis, M. H.; Balfanz, R. W.; and Fox, J. H. 2019. An efficacy study of a ninth-grade early warning indicator intervention. *Journal of Research on Educational Effectiveness*, 12(3): 363–390.

Mate, A.; Madaan, L.; Taneja, A.; Madhiwalla, N.; Verma, S.; Singh, G.; Hegde, A.; Varakantham, P.; and Tambe, M.

2022. Field study in deploying restless multi-armed bandits: Assisting non-profits in improving maternal and child health. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI '22)*, 12017–12025.

Mate, A.; Wilder, B.; Taneja, A.; and Tambe, M. 2023. Improved Policy Evaluation for Randomized Trials of Algorithmic Resource Allocation. In *Proceedings of the 40th International Conference on Machine Learning (ICML '23)*, 24198–24213.

Murthy, N.; Chandrasekharan, S.; Prakash, M. P.; Ganju, A.; Peter, J.; Kaonga, N.; and Mechael, P. 2020. Effects of an mHealth voice message service (mMitra) on maternal health knowledge and practices of low-income women in India: findings from a pseudo-randomized controlled trial. *BMC public health*, 20: 1–10.

Perdomo, J. C. 2024. The Relative Value of Prediction in Algorithmic Decision Making. In *Proceedings of the Forty-first International Conference on Machine Learning (ICML '24)*.

Senn, S. S. 2008. *Statistical issues in drug development*, volume 69. John Wiley & Sons.

Shah, S.; Suggala, A. S.; Tambe, M.; and Taneja, A. 2024. Efficient Public Health Intervention Planning Using Decomposition-Based Decision-focused Learning. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems, (AAMAS '24)*, 1701–1709.

Shirali, A.; Abebe, R.; and Hardt, M. 2024. Allocation Requires Prediction Only if Inequality Is Low. In *Proceedings of the Forty-first International Conference on Machine Learning (ICML '24)*.

van der Vaart, A.; and Wellner, J. A. 2023. Empirical processes. In *Weak Convergence and Empirical Processes: With Applications to Statistics*, 127–384. Springer.

Verma, S.; Mate, A.; Wang, K.; Madhiwalla, N.; Hegde, A.; Taneja, A.; and Tambe, M. 2023a. Restless Multi-Armed Bandits for Maternal and Child Health: Results from Decision-Focused Learning. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS '23)*, 1312–1320.

Verma, S.; Singh, G.; Mate, A.; Verma, P.; Gorantla, S.; Madhiwalla, N.; Hegde, A.; Thakkar, D.; Jain, M.; Tambe, M.; et al. 2023b. Expanding impact of mobile health programs: SAHELI for maternal and child care. *AI Magazine*, 44(4): 363–376.

Verma, S.; Zhao, Y.; Shah, S.; Boehmer, N.; Taneja, A.; and Tambe, M. 2024. Group Fairness in Predict-Then-Optimize Settings for Restless Bandits. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence (UAI '24)*, 3448–3469.

Wager, S.; and Athey, S. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523): 1228–1242.

Wang, K.; Verma, S.; Mate, A.; Shah, S.; Taneja, A.; Madhiwalla, N.; Hegde, A.; and Tambe, M. 2023. Scalable decision-focused learning in restless multi-armed bandits with application to maternal and child health. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI '23)*, volume 37, 12138–12146.

Weber, R. R.; and Weiss, G. 1990. On an index policy for restless bandits. *Journal of Applied Probability*, 27(3): 637–648.

Yeter, B.; Garbatov, Y.; and Soares, C. G. 2020. Risk-based maintenance planning of offshore wind turbine farms. *Reliability Engineering & System Safety*, 202: 107062.